Biometrics Day

Behavioural Authentication by Orange

Motivations, techniques, privacy, and next steps

July 07, 2017

J-J Schwartzmann
Content

1. Motivations
2. Techniques
3. Privacy protection
4. Next steps
How authenticate humans?

• in real life

• Bob meets Alice

• He knows she is really Alice thanks to some physical and behavioural features such as:
  • gait
  • suit
  • voice
  • hair
  • face

• Bob doesn't challenge Alice for proofs: no need to exchange secret information
How authenticate humans?

• in the digital world

• Alice wants to access her online bank

• The bank server asks her to provide authentication proofs such as:
  • password (knowledge factor)
  • OTP (ownership factor)
  • fingerprint (biometric factor)
Behavioural authentication: a paradigm shift

• why the bank server should not recognize Alice as Bob does it?

• the continuous collection of multiple behavioural factors makes natural authentication possible:
  • transparent and seamless
  • secure due to the multiplicity of behavioural data [1]
  • continuous instead of static (e.g. password on web) or periodic (e.g. Windows session): permanent trust level
  • adaptive authentication: trust level can be just maintained to the services requirements
  • but behavioural data are sensitive: what about privacy?
Continuous trust score

toward an adaptive authentication

- A service provider requires different levels of confidence to authorize the entity to perform different actions.
- Authentication is a continuous process, depending not only on authentication factors, but also on context elements such as time.
- Global confidence, resulting from a set of different elements, should not be described as a discrete value.
- 4 discrete Levels of Assurance based on a set of requirements are not sufficient to describe the real world: a fine measurement in real time of the level of confidence would allow to better align authentication processes on requirements of relying parties.

Orange’s Scientific Council 2014

Google I/O 2015
Our recent work results

- Julien Hatin recruited as doctoral student in 2014
- thesis topic: "Trust evaluation in an authentication process"
- major deliverable results:
  - "A Continuous LoA Compliant Trust Evaluation Method, ICISSP 2016: this paper propose a model based on the Dempster-Shafer theory to merge continuous authentication system with more traditional static authentication scheme and to assign a continuous trust level, compliant with the LoA [2]
  - 2016 Orange's Research Exhibition: presentation of our proof of concept named "Behavioural Authentication" based on swipe gestures, GPS-call, and password
  - "Privacy Preserving Transparent Mobile Authentication", ICISSP 2017: we propose a solution to address privacy issues using the BioHashing algorithm on behavioural information extracted from a mobile phone [3]
1. Motivations

2. Techniques

3. Privacy protection

4. Next steps
Behavioural modalities

- Google's proposal
  - swipe gestures
  - gait recognition
  - dynamic keystroke
  - voice verification
  - face recognition

- fusion (simple AND in this case)
  $\Rightarrow$ 10x more accurate than fingerprint authentication (according to Google)
Unimodal biometric processing chain

- **Sensor**: fingerprint sensor/camera, keyboard/touchscreen, accelerometer/gyroscope, microphone, GPS/GSM antenna, software...
- **Features Extractor**: image processing, signal processing, statistical analysis...
- **Matcher**: Hamming Distance, Dynamic Time Warping, k-Nearest-Neighbours, Support Vector Machine, Neural Networks, Gaussian Mixture Model, Naive Bayes...
- **Database**: template features
- **Rank/Decision**: accepted/rejected OR trust level

Data source:
- raw data
- features
- distance or score
From data source to features

- hardware or software sensors to capture data sources
  - dedicated: fingerprint sensor (Apple TouchID)
  - general purpose: touchscreen, camera, microphone, accelerometer, etc.
  - software based

- preprocessing raw data
  - noise reduction
  - pre-emphasis
  - filtering

- features extractor returns a feature vector as result
  - image processing (morphological biometrics: face, iris, fingerprint)
  - signal processing (voice, ECG)
  - statistical analysis (swipe gestures, dynamic keystroke, call and usage habits)
Matching and decision: a data mining problem

• matcher
  • returns a distance (integer or real) or similarity score in [0,1]
  • computing distance: hamming, DTW
  • using single or multiple two-class classifiers: SVM, kNN, GMM, HMM, etc. [4]

• from the classifier output to decision
  • comparison with acceptance threshold: return boolean decision,
  • or computing a discrete or continuous trust level
Fusion in multimodal systems

- information fusion can be done at different levels
  - rank/decision-level
  - score-level
  - feature-level

- rank/decision-level fusion
  - boolean functions
  - fuzzy logic
  - bayesian networks
  - theory of belief functions (subjective logic, Dempster-Shafer theory [2])
Fusion at rank/decision-level

Data source

Sensor

Raw data

Features extractor

Features

Matcher

Distance or score

Rank/decision

Rank/decision fusion

Accepted/rejected (or trust level)

Features

Template database

Sensor

Raw data

Features extractor

Features

Matcher

Distance or score

Rank/decision
Fusion at score-level

sensor
data source

raw data

features extractor

features

distance or score

matcher

score fusion

raw data

features extractor

distance or score

matcher

features

data source

template database

accepted/rejected
(or trust level)

decision
What experts do we need?

- continuous authentication is a wide domain relying on various branches of applied mathematics
  - signal and image processing
  - data mining and machine learning: requires almost as many experts as there are biometric modalities ;-) 
  - decision theory
- development of advanced prototypes need senior developers
- acceptability issues need social science experts
Future directions and scientific locks

- improvement of features extraction so that most appropriate features are selected
- the biometric data at enrollment time may have different characteristics than those presented during authentication. Domain adaptation and transfer learning techniques can be used to deal with this changing distribution problem
- lack of large data sets: most behavioural authentication techniques have been evaluated on small datasets. They should be evaluated on large-scale data sets containing millions of samples
- spoof, mimic, and replay attacks
- revocability of biometric data and privacy issues
- acceptability issues
Content

1. Motivations

2. Techniques

3. Privacy protection

4. Next steps
How to protect biometric data?

- **cryptographic hash functions** (as used for passwords protection)
  - the diffusion property (avalanche effect) make those functions unusable because of the variability of biometric data

- **template encryption**
  - encryption does not ensure privacy protection
  - decryption needed before matching implies a potential security lack
  - homomorphic encryption is today too slow (especially on mobile devices)

- **secure computing**
  - Match-On-Card (MOC), especially on UICC [5]
  - TEE (Apple TouchID)
  - suitable only for on-device authentication
Cancelable biometrics

• concept introduced in 2001 by Ratha et al. [6]

• transforming features such as transformed data should respect the following properties [7]
  
  • revocability: the data should be easily revoked in case of compromise
  
  • non-invertibility: one should not be able to obtain information on the raw features from the transformed data (pre-image resistance)
  
  • performance: the transformation should not alter the accuracy of the biometric verification system
  
  • diversity: on the basis of the same features, one should be able to generate different transformed data. The cross-checking of these different data should not make it possible to retrieve information on the raw features
Cancelable biometrics scheme [8]

- the original features are not stored
- only the BioCode is stored
- it is not possible to compute the pattern or retrieve the original image given the BioCode
- a new BioCode can be re-generated with another key in case of compromise
- similarity is preserved [9] so matching is performed by computing a Hamming distance

\[
\begin{bmatrix}
  b_1 \\
  b_2 \\
  \vdots \\
  b_m
\end{bmatrix}
\]

\[ m \leq n \]

\[ b_i \in \{0,1\} \]
BioHashing algorithm [8]

- consists in projecting the $n$ original biometric features on a $m$ dimension basis ($m \leq n$). The final result is binarized and stored as the BioCode.

\[
\begin{bmatrix}
    r_{1,1} & \cdots & r_{1,m} \\
    \vdots & \ddots & \vdots \\
    r_{n,1} & \cdots & r_{n,m}
\end{bmatrix}
\]

Gram-Schmidt orthonormalisation

\[
\begin{bmatrix}
    o_{1,1} & \cdots & o_{1,m} \\
    \vdots & \ddots & \vdots \\
    o_{n,1} & \cdots & o_{n,m}
\end{bmatrix}
\]

random generator

$m \times n$ matrix

raw biometric data

\[
\begin{bmatrix}
    f_1 & \cdots & f_n
\end{bmatrix}
\times
\begin{bmatrix}
    o_{1,1} & \cdots & o_{1,m} \\
    \vdots & \ddots & \vdots \\
    o_{n,1} & \cdots & o_{n,m}
\end{bmatrix} =
\begin{bmatrix}
    p_1 & \cdots & p_m
\end{bmatrix}
\]

binarization

\[
\begin{bmatrix}
    B_1 & \cdots & B_m
\end{bmatrix}, B_i \in \{0,1\}
\]

Binarization

BioCode
BioHashing applied to located call data

- call habits combined with location are proved to be relevant as behavioural authentication data [10]
- data used are: phone number of the callee, latitude and longitude of the cell
- such data are very sensitive in terms of privacy so they have to protected
Content

1. Motivations

2. Techniques

3. Privacy protection

4. Next steps
Work in progress

- release a pre-industrial framework by end of 2017
- Android clients collects and scrambles raw datas to extract behavioural features
- several retro-controlled micro-services compute modalities scores
- fusion of scores get a trust level which is a new authentication claim usable by a standard Identity Provider
Assessment of the technology

- experiments at scale (end 2017–2018)
  - a data collection with Cité des Télécoms (Pleumeur Bodou – 22)
  - an ANR project to test the framework in vivo with IRT Bcom
  - potential Orange internal anticipation projects

⇒ be the first to demonstrate accuracy of behavioural biometrics on large datasets

- measuring users' acceptance (end 2017–2020)
  - a new research activity thanks to information and communication science
  - a new thesis topic: "Building institutional trust: application to digital identity solutions"
new thesis topic for 2017–2020

• "behavioural authentication in an ubiquitous digital environment"

• users operate across multiple devices, including desktop PCs, laptops, tablets and smartphones. As a consequence, they can regularly find themselves having a variety of devices open concurrently, so there is a resultant need to repeatedly authenticate

• the purpose of the thesis is to provide an "authentication aura" to the users, using behavioural data from those various sources, including wearable devices
Questions ?
Appendices
## Levels of Assurance


<table>
<thead>
<tr>
<th>LoA</th>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Low</td>
<td>There is no specific requirement for the authentication mechanism used; only that it provides some minimal assurance. A wide range of available technologies, including the credentials associated with higher LoAs, can satisfy the authentication requirements for this LoA. This level does not require use of cryptographic methods.</td>
</tr>
<tr>
<td>2</td>
<td>Medium</td>
<td>Single-factor authentication is acceptable. Successful authentication shall be dependent upon the entity proving, through a secure authentication protocol, that the entity has control of the credential. Controls shall be in place to reduce the effectiveness of eavesdropper and online guessing attacks. Controls shall be in place to protect against attacks on stored credentials.</td>
</tr>
<tr>
<td>3</td>
<td>High</td>
<td>This LoA shall employ multi-factor authentication. Identity proofing procedures shall be dependent upon verification of identity information. Any secret information exchanged in authentication protocols shall be cryptographically protected. There are no requirements concerning the generation or storage of credentials; they may be stored or generated in general purpose computers or special purpose hardware.</td>
</tr>
<tr>
<td>4</td>
<td>Very high</td>
<td>LoA4 provides the highest level of entity authentication assurance defined by this Recommendation</td>
</tr>
</tbody>
</table>
References (1/2)


References (2/2)